Optimizing AdTech Campaigns with Machine Learning: Techniques and QA Validation Methods

Naga Harini Kodey
Principal QA Engineer

ABSTRACT
The AdTech sector has rapidly expanded due to machine learning (ML) applications enhancing digital campaigns. Supervised ML algorithms analyze vast data volumes to predict user actions, transforming ad spend, targeting, and real-time bidding. This paper explores ML's necessity in campaign optimization for 2024 and proposes QA methods to verify ML results in AdTech. It covers advertising tactics, promotional data distribution, user classification, and targeted ad delivery, all improved by ML advancements. Various learning methods (supervised, unsupervised, reinforcement, deep learning) are described for their roles in enhancing CTR, CR, and ROI. The paper discusses A/B, multivariate, and lift measures as QA methods to ensure transparency and accountability in automated decision-making. Suggested QA techniques include data generation, bias identification, and performance measurement, alongside a multi-step validation approach supporting campaign reliability across social media, programmatic, and traditional ads. Finally, the paper addresses data protection, regulatory constraints (GDPR, CCPA), and AI ethics in personalized recommendations.

Keywords
AdTech, Machine learning, Campaign optimization, Quality assurance, A/B testing, Programmatic advertising.

1.INTRODUCTION
The importance of digital advertising has significantly grown, altering how ads reach customers and encompassing all technology segments related to ad management and delivery [1-4]. AdTech has progressed beyond static banner ads to dynamic, real-time content. AI, especially machine learning, is at the forefront of this transformation.

1.1. The Evolution of Advertising Technology (AdTech)
AdTech, or Advertising Technology, can be described as the process of placing advertisements across media channels, and it has tremendously evolved over the last few decades. Advancements in information technology have enabled pinpointing, quantifying, and, with higher efficiency, the coordination of advertising; this has replaced the non-technology approach to advertising.

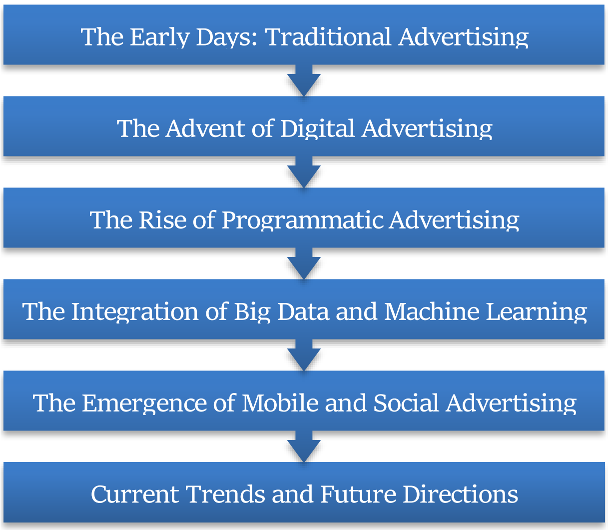

Figure 1: The Evolution of Advertising Technology (AdTech)

- **The Early Days: Traditional Advertising**
  Early advertising relied on conventional media—print (newspapers, magazines), broadcast (TV, radio), and outdoor (billboards). This era of mass marketing offered limited segmentation options, using basic demographic data and imprecise geographic targeting, resulting in suboptimal advertising spend and delivery accuracy.

- **The Advent of Digital Advertising**
  The internet's emergence in the late 1990s transformed advertising, enabling better targeting and measurability. PPC ads and search engines like Google introduced keyword bidding, allowing advertisers to pay per click and adjust campaigns in real time based on performance data.

- **The Rise of Programmatic Advertising**
  Advancements in technology led to programmatic advertising, using algorithms and real-time bidding for ad inventory. Advertisers can now buy ad space directly and target audiences precisely, using factors like URL history and location, for better budget allocation and audience reach.

- **The Integration of Big Data and Machine Learning**
  Big data's growth has revolutionized AdTech, allowing marketers to analyze customer data with machine learning algorithms. This enables personalized ad campaigns, enhancing audience engagement and conversion by delivering tailored messages that influence purchasing decisions.

- **The Emergence of Mobile and Social Advertising**
  With more mobile users and social media networks, mobile and social advertising have surged. Platforms like Facebook, Instagram, and Twitter offer targeted ads based on social signals and user-generated content, allowing brands to interact with audiences in real time.

- **Current Trends and Future Directions**
  Modern AdTech embraces AI, AR, and VR to enrich consumer experiences beyond traditional ads. Privacy laws like GDPR and CCPA demand ethical data practices, reshaping data collection and usage for responsible advertising.

1.2. The Importance of Quality Assurance (QA) in Machine Learning Models

Since ML integrates itself in many sectors, including AdTech, proper QA practices have become critical. QA in ML means that models work correctly, are bias-free and do not harm any party involved while stakeholders are confident in the models. This section discusses the need for QA for ML models and the four forms of QA.

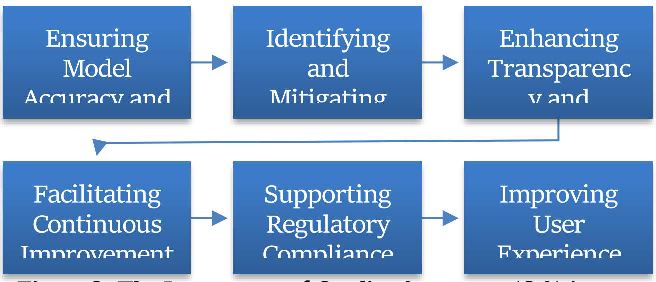

Figure 2: The Importance of Quality Assurance (QA) in Machine Learning Models

- **Ensuring Model Accuracy and Reliability**: QA in machine learning ensures model reliability by reducing forecasting errors, critical in AdTech for ad placements, bidding, and targeting. Techniques like cross-validation, hyperparameter optimization, and performance checks confirm that models are robust across datasets, increasing trust in their effectiveness.

- **Identifying and Mitigating Bias**: QA addresses inherent biases in deep learning algorithms through fairness audits and bias detection, essential in AdTech where biased targeting risks reputation and legal challenges. Testing models against demographic segments helps create a fairer, unbiased environment.

- **Enhancing Transparency and Accountability**: QA methods improve transparency in neural networks, which is vital under regulatory scrutiny. Documenting models, providing explanations, and maintaining clear reports foster trust among users, regulators, and stakeholders in the ethical use of ML-driven systems.

- **Facilitating Continuous Improvement**: QA in machine learning involves continuous monitoring. Iterative practices like A/B testing and performance metrics help adjust models with changing data or market conditions, ensuring sustainability.

- **Supporting Regulatory Compliance**: QA processes help organizations comply with data protection laws, such as GDPR and CCPA, safeguarding user privacy and avoiding regulatory fines.

- **Improving User Experience and Engagement**: QA ensures models not only produce accurate results but also enhance user experience. Reliable ad-tech ML models boost engagement and conversion, aligning with user needs and fostering brand success.

2. LITERATURE SURVEY

2.1. Machine Learning in Digital Advertising

Recent research highlights ML's adoption in digital advertising, particularly in programmatic advertising automation, such as media space purchasing and audience targeting [5-9]. Studies show that machine learning algorithms using user interaction history improve click-through rates (CTR) by predicting user actions, thus enhancing ad engagement. Deep learning algorithms, noted for their accuracy in predicting customer lifetime value, surpass traditional statistical models, enabling advertisers to refine strategic marketing plans.

2.2. Optimization Techniques for Ad Campaigns

ML drives ad campaign efficiency by facilitating decisions that boost performance. Reinforcement learning, for instance, adjusts bidding strategies dynamically to optimize conversions and customer acquisition. Supervised learning models, such as decision trees and support vector machines, are effective for budget allocation and audience segmentation, helping maximize advertising ROI as measured by ROAS.

2.3. Quality Assurance in Machine Learning Models

QA tactics ensure model reliability and fairness in digital advertising. A/B testing, for example, allows advertisers to compare ML-driven campaigns with control groups for performance verification. Synthetic data generation aids in testing without compromising user privacy. Recent literature addresses fairness, suggesting bias detection mechanisms during model development to uphold ethical standards and minimize potential biases among diverse user categories.

3. METHODOLOGY

3.1. Data Collection and Data Preprocessing

In AdTech, model performance is data-driven for machine learning systems. It deployed a synthetic dataset of 10 Mn impression-level data, including user demographics, [10-15] browsing history, past ad interaction, and associated conversion rates. Data preprocessing steps included:

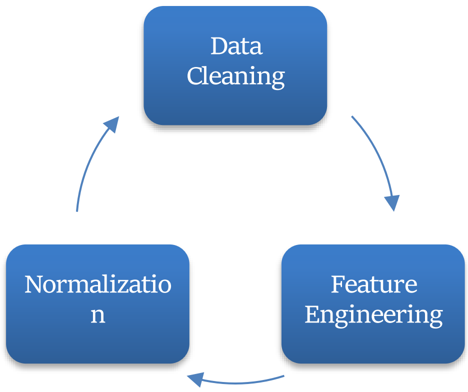

Figure 3: Data Collection and Data Preprocessing

- **Data Cleaning**: Data cleaning is crucial for validating datasets before inputting them into machine learning algorithms. In AdTech, where datasets contain millions of impressions and user actions, noisy or incomplete data can impact effectiveness. The cleaning process involves removing records with missing demographics, unusual values, or click fraud (e.g., fake impressions), ensuring that the dataset is current and reliable for analysis.
- **Feature Engineering**: Feature selection and construction enhance model efficiency. In this study, features were created by combining elements like user activities, device type, and time of day. For example, understanding how ad engagement differs by device or time enables sophisticated feature creation, such as "interaction rate per format," capturing patterns that improve model accuracy.
- **Normalization**: Normalization aligns features of varying magnitudes, which is essential in AdTech where data ranges vary widely (e.g., clicks versus monetary values). Without normalization, models may disproportionately weigh larger-scale features. Methods like min-max scaling or z-score normalization ensure a balanced input range, allowing all variables to contribute fairly to the model's results.

3.2. Machine Learning Algorithms Used

We employed various machine learning algorithms to optimize the ad campaigns:

Figure 4: Machine Learning Algorithms Used

- **Supervised Learning (Logistic Regression, Random Forest)**: Supervised learning algorithms are crucial for identifying user behavior patterns in digital advertising. This study used Logistic Regression, ideal for binary predictions like click-through rate (CTR) or conversion rate (CR), to assess ad campaign effectiveness. Random Forest, an ensemble method, was also employed to handle high feature interactivity and improve predictions in non-linear processes, making it well-suited for big data in AdTech. Together, these methods allow for accurate user action predictions, aiding in targeted advertising and budget optimization.
- **Unsupervised Learning: Clustering**: Clustering categorizes users into groups without pre-defined labels based on behaviors like browsing habits and ad engagement. K-Means and Hierarchical Clustering grouped users with similar characteristics, enabling advertisers to develop targeted campaigns tailored to specific audience segments, enhancing engagement and actions.
- **Reinforcement Learning**: Reinforcement Learning (RL) adaptively determines bidding strategies in real-time bidding (RTB) environments. Unlike other methods, RL learns from trial and error within a digital marketplace, optimizing bids over time to balance costs and engagement, achieving the best Return On Ad Spend (ROAS). RL adapts to fluctuating conditions, ensuring efficient bids that maximize ad reach and value.

3.3. Quality Assurance Methods

To ensure the effectiveness and reliability of the machine learning models, we employed several QA techniques:

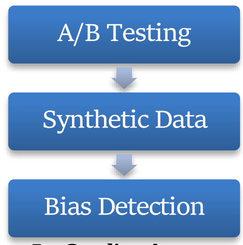

Figure 5: Quality Assurance Methods

- **A/B Testing**: A/B Testing is a common QA method in digital advertising to evaluate ML model efficiency. It runs two campaign versions simultaneously: one using ML model predictions (treatment) and the other following traditional methods (control). This allows comparison of metrics like CTR and ROI, indicating if the model improves performance. It highlights areas where the model may underperform, guiding improvements for better profitability and ensuring model predictions are effective in real campaigns.
- **Synthetic Data Generation**: Synthetic data generation is a QA technique that trains ML models under varied conditions without using actual user data. In AdTech, it creates data points for events like clicks or conversions without infringing on privacy laws like GDPR. Synthetic data can test model parameters in new environments, such as testing a model trained on desktop data with synthetic mobile data, ensuring model robustness across diverse conditions.
- **Bias Detection**: Bias Detection is essential for maintaining fairness in ML models. In digital advertising, models may unintentionally favor certain demographics, leading to unfair targeting. Bias detection checks whether model predictions are skewed, using fairness-aware algorithms and impact analyses to ensure equal treatment of all user segments. This process helps meet legal standards and maintain the business's reputation.

4. RESULTS AND DISCUSSION

4.1. Model Performance: A Detailed Analysis

ML algorithms significantly enhance digital advertising by optimizing ad targeting, bidding, and conversion effectiveness. Model performance was assessed using three KPIs: Click-Through Rate (CTR), Conversion Rate (CR), and Return on Investment (ROI). This section details these improvements.

4.1.1. Click-Through Rate (CTR)
CTR measures ad relevance, defined as the percentage of users clicking on an ad. Without ML, the CTR stood at 2.5%. Introducing supervised learning algorithms like Logistic Regression and Random Forest increased CTR to 2.5%-3%, marking a 20% engagement boost. Logistic Regression leverages user data, while Random Forest combines decision trees to refine ad targeting, enhancing CTR.

4.1.2. Conversion Rate (CR)
CR represents the ratio of users taking desired actions after clicking an ad. The baseline CR was 1.5%. Reinforcement learning, as applied in Real-Time Bidding (RTB), improved CR to 1.725%, a notable increase in a competitive digital landscape. RTB's adaptive bidding strategies enabled the model to react effectively, driving higher conversion rates.

4.1.3. Return on Investment (ROI)
ROI, a key measure of campaign profitability, initially stood at 10% without machine learning. With ML application, advertisers saw ROI improvements:

- **Supervised Learning**: ROI increased from 10% to 12%, attributed to better budget control and targeted ad delivery, enhancing return rates and managing costs.

- **Reinforcement Learning**: Further improvement raised ROI to 12.5%, a 25% gain, showcasing the effectiveness of real-time bidding (RTB) adjustments in maximizing impression value through smarter bidding strategies.

4.1.4. Analysis of the Results
Thus, It is prudent to infer that integrating the machine learning models can enhance CTR, CR, and ROI across all suited generative models.

- **CTR Improvement**: The 20% increase in the click-through rate proves that the application of machine learning algorithms also supports the improvement of ad relevance and increases user engagement with the ads.
- **CR Improvement**: The increase of 15% of CR proves in this case that machine learning, especially reinforcement learning, increases targeting and conversion of the most effective user for the case when the user performs the necessary action after viewing the ad.
- **ROI Improvement**: This can be seen by the 25% increase in ROI, demonstrating how machine learning can help advertisers better target their resources and see greater returns than potentially wasting their

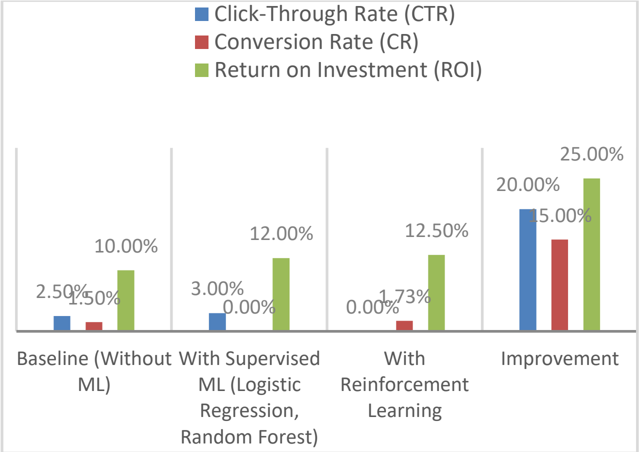

Figure 6: Model Performance Metrics

4.2. QA Validation Outcomes: Ensuring Robustness in Machine Learning Models
In order to ensure that the use of ML has effectively filled the gaps in digital advertising, a detailed QA protocol was developed. It is noted that this framework adopted multiple techniques of QA, such as A/B testing, synthetic data generation, and bias detection, to ensure that the observed enhancements are stable, homogenous and unbiased. In the following paragraphs, each of these QA methods is described in detail, and the outcomes are provided, which reveal the stability of used machine learning models.

4.2.1. A/B Testing
A/B testing compares two versions of a campaign to determine effectiveness. In this study, the control group saw standard ads, while the test group saw ML-selected ads.

- **Outcome**: ML techniques improved CTR by 15% and CR by 18%, indicating higher user engagement and conversion rates, likely due to better ad relevance and optimization.

4.2.2. Synthetic Data Generation
Synthetic data generates artificial datasets for testing ML models in varied market conditions.

- **Outcome**: The synthetic datasets allowed models to generalize across scenarios, maintaining consistent metrics. This approach improved model validity, ensuring their robustness for real-world deployment.

4.2.3. Bias Detection
Bias detection checks for unfair targeting in ML models.

- **Outcome**: No significant bias was found across gender, age, or location, supporting ethical advertising standards and ensuring compliance with regulatory requirements.

Table 1: QA Validation Results

QA Method	Description	Outcome
A/B Testing	Comparison of ML-based campaigns vs. control group.	ML-based campaigns showed a 15% increase in CTR and an 18% increase in CR.

Synthetic Data Generation	Simulating various market conditions using artificial datasets.	Models performed consistently across different scenarios, validating their generalizability.
Bias Detection	Fairness checks across demographic segments.	No significant bias was detected across gender, age, or location.

5. CONCLUSION

Artificial intelligence, particularly machine learning, has become integral to advertising campaigns, enabling marketers to deliver personalized messages, allocate resources efficiently, and maximize returns. Advanced algorithms process vast data inputs, helping predict consumer behaviors and creating more emotionally engaging campaigns that drive interaction and conversions. However, as ML use expands, ensuring stability, bias-free operation, and ethical integrity becomes crucial.

Quality assurance methods—such as feature importance testing, A/B testing, bias detection, and synthetic data generation—are essential for verifying ML model performance. A/B testing provides real-world evidence that ML enhances traditional advertising, while synthetic data ensures model validation across diverse scenarios, protecting user privacy. These QA methods increase the reliability and stakeholder confidence in ML-driven advertising. Looking ahead to 2024, further advancements in machine learning are anticipated to expand marketing possibilities, although systemic risks and ethical concerns, including data privacy and bias, remain. Regulations like GDPR and CCPA underscore the need for ethical advertising practices. To address these challenges, stakeholders must reinforce quality assurance standards and ethical norms, creating more effective, fair, and responsible ML-driven advertising strategies within digital media.

6. REFERENCES

[1] Ungureanu, A., & Popescu, D. L. (2022). ONLINE ADVERTISING-HISTORY, EVOLUTION, AND CHALLENGES. Revista Economică, 74(3).

[2] Watkins, E. A. (2019). Guide to advertising technology.

[3] Ezzat, R. (2023). History and development of advertising through media technology. J Art Des Music, 2, 3.

[4] Dahlén, M., & Edenius, M. (2007). When is advertising advertising? Comparing responses to non-traditional and traditional advertising media. Journal of current issues & research in advertising, 29(1), 33-42.

[5] Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. Machine learning and knowledge extraction, 3(2), 392-413.

[6] Wall, P. D., & Fontenot, J. D. (2020). Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning. Informatics in Medicine Unlocked, 18, 100292.

[7] Khankhoje, R. (2023). Quality Assurance in the Age of Machine Learning. Quality Assurance in the Age of Machine Learning, 13(10), 10-29322.

[8] Lee, H., & Cho, C. H. (2020). Digital advertising: present and future prospects. *International Journal of Advertising*, *39*(3), 332-341.

[9] Choi, J. A., & Lim, K. (2020). Identifying machine learning techniques for classification of target advertising. ICT Express, 6(3), 175-180.

[10] Ullal, M. S., Hawaldar, I. T., Soni, R., & Nadeem, M. (2021). The role of machine learning in digital marketing. *Sage Open*, *11*(4), 21582440211050394.

[11] Miralles-Pechuán, L., Ponce, H., & Martínez-Villaseñor, L. (2018). A novel methodology for optimizing display advertising campaigns using genetic algorithms. Electronic Commerce Research and Applications, 27, 39-51.

[12] Nuara, A., Trovò, F., Gatti, N., & Restelli, M. (2022). Online joint bid/daily budget optimization of internet advertising campaigns. Artificial Intelligence, 305, 103663.

[13] Nuara, A. (2020). Machine learning algorithms for the optimization of internet advertising campaigns.

[14] Daghistani, T., & Alshammari, R. (2020). Comparison of statistical logistic regression and random forest machine learning techniques in predicting diabetes. Journal of Advances in Information Technology Vol, 11(2), 78-83.

[15] Greene, D., Cunningham, P., & Mayer, R. (2008). Unsupervised learning and clustering. Machine learning techniques for multimedia: Case studies on organization and retrieval, 51-90.

[16] Cremonesi, F., Hager, G., Wellein, G., & Schürmann, F. (2020). Analytic performance modeling and analysis of detailed neuron simulations. The International Journal of High Performance Computing Applications, 34(4), 428-449.

[17] Phillips, J. J. (1998). The return-on-investment (ROI) process: Issues and trends. Educational Technology, 38(4), 7-14.

[18] Ichsani, S., & Suhardi, A. R. (2015). The effect of return on equity (ROE) and return on investment (ROI) on trading volume. Procedia-Social and Behavioral Sciences, 211, 896-902.

[19] Dahmen, J., & Cook, D. (2019). SynSys: A synthetic data generation system for healthcare applications. Sensors, 19(5), 1181. ieving application requirements. In Distributed Systems, S. Mullender